



ISSN: 1646-8929

IET Working Papers Series

No. [WPS03/2010](#)

João Bolinhas

(email: joaobolinhas@gmail.com)

Ricardo Neves

(email: ricardonevesfct@gmail.com)

O futuro da Web Semântica e sua importância nas organizações
--

IET

Research Centre on Enterprise and Work Innovation

Centro de Investigação em Inovação Empresarial e do Trabalho

Faculdade de Ciências e Tecnologia

Universidade Nova de Lisboa

Monte de Caparica

Portugal

O futuro da Web Semântica e sua importância nas organizações¹

[The future of semantic web and it's importance at organisations]

[entregue para publicação: Quinta-feira, 25 de Março de 2010]

Índice

Resumo.....	4
Abstract.....	5
1.Introdução.....	6
2.Web Semântica.....	8
2.1 O que é a Web Semântica?.....	8
2.2 Tecnologias de Base	8
2.2.1 Resource Description Framework (RDF).....	8
2.2.2. Ontology Web Language (OWL).....	9
2.3 Ontologias.....	9
2.4 Taxonomias.....	9
3.A Web Semântica nas Organizações.....	10
3.1 A importância da Web Semântica nas Organizações.....	10
3.2 Desafios	11
3.3 Casos de Uso.....	11
3.3.1 Melhorar a Eficácia dos Serviços de Recrutamento.....	11
3.3.2 Melhorar a manufatura ágil.....	13

1 Trabalho realizado sob orientação do Prof. António Brandão Moniz para a disciplina “Factores Sociais da Inovação” do Mestrado Engenharia Informática realizado na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (Portugal).

3.3.3 Integração de dados científicos heterogéneos	14
3.3.4 Optimização da procura e navegação em uma empresa	16
3.3.5 Áreas Adicionais da Web Semântica.....	17
4.Projectos de Investigação na Web Semântica.....	18
4.1 Semantic Oriented Multimedia Indexation and Retrieval	19
4.2 Large scale semantic computing semantic Web technologies distributed reasoning probabilistic reasoning web-scale inference information retrieval (LARKC).....	20
4.3 Knowledge in a Wiki (KIWI).....	21
4.4 Collaborative annotation of a large biomedical corpus (CALBC).....	22
4.5 OntoWiki semantic collaboration for enterprise knowledge management, e-learning and e- tourism (ONTOWIKI).....	23
4.6 CASPOE – Caracterização Semântica e Pragmática de Objectos Educativos.....	24
4.7 SemanticShare – Ferramentas de Recursos para o Processamento Semântico.....	26
4.8 Informação Estatística.....	27
5.Conclusão.....	31
6.Bibliografia.....	32

Resumo

Actualmente a Web Semântica e as potencialidades da sua tecnologia extravasam completamente a visão de *Tim Berners-Lee* apresentada em 1999, no qual o significado (semântica) das informações e serviços na Web é definido, tornando possível a sua compreensão pelos computadores, de modo a obter uma experiência de utilização mais rica. Hoje em dia, a comunidade científica e académica perceberam que a Web Semântica pode ser uma importante ferramenta para facilitar a integração e a interoperabilidade dos processos internos e externos empresariais. Permitindo assim às empresas, num mundo totalmente globalizado onde se lidam com grandes volumes de informação, continuar a atingir os objectivos de negócio em face da mudança e da crescente dependência do uso efectivo de informação. Este trabalho focará a importância que esta tecnologia vem adquirindo no melhoramento dos processos das organizações, materializada em alguns casos de usos possíveis. Para além disso, será efectuada uma visão geral sobre os projectos de investigação que decorrem, de modo a que se possam compreender as tendências futuras desta tecnologia.

Palavras-chave: Web Semântica nas Organizações; web semântica; manufactura ágil.

Abstract

Currently, the Semantic Web and the potential of their technology completely go beyond the vision of Tim Berners-Lee's 1999, in which the meaning (semantics) of information and services on the web is defined, making possible their understanding for computers, to obtain a richer user experience. Today, the scientific and academic community realized that the Semantic Web can be an important tool to facilitate integration and interoperability of the internal and external business processes. Allowing the companies, in a fully globalized world which deal with large volumes of information, continue to achieve business goals in the face of change and the growing dependence of the effective use of information. This work will focus on the importance of this technology has been gaining in improving the processes organizations, embodied in some possible use cases. In addition, there will be an overview of the research projects taking place, in order to understand the future trends of this technology.

Keywords : semantic web on organizations; semantic web; agile manufacturing.

JEL codes: M11; M15; O30

1.Introdução

A ideia original da Web Semântica foi a de fornecer uma descrição dos dados e documentos existentes na Web, que pudesse ser interpretada pelas máquinas, a fim de melhorar a pesquisa e o uso dos dados. A *World Wide Web* era, e na maioria dos casos continua a ser, um vasto conjunto de páginas Web dinamicamente ligadas. As páginas são escritas em HTML (*Hyper Text Markup Language*), uma linguagem que é bastante útil para a publicação de informação destinada exclusivamente ao consumo humano. Os seres humanos conseguem ler páginas Web e compreendê-las correctamente, porém o significado inerente não está disponível numa forma que permita a interpretação por parte dos computadores.

A Web Semântica visa definir maneiras de permitir que as informações da Web possam ser utilizadas pelos computadores, não só para fins de visualização, mas também para a interoperabilidade e integração entre sistemas e aplicações. Uma forma de permitir esse processamento automatizado é fornecer informações de tal forma, que os computadores possam entender. Para dar significado à informação existente na rede, novos standards e linguagens estão a ser investigadas e desenvolvidas. Exemplos bem conhecidos incluem: *Resource Description Framework* (RDF)² e *Web Ontology Language* (OWL)³. A informação descritiva disponível pela utilização destas linguagens, permite caracterizar individualmente e de uma forma precisa todo o tipo de recursos existentes na Web e respectivas relações entre eles.

Hoje, a Web Semântica não é apenas sobre o aumento da expressividade da informação existente na rede. Os académicos e as organizações perceberam que a Web Semântica pode facilitar a integração e a interoperabilidade dos processos internos e externos de negócios, bem como permitir a criação de infra-estruturas globais para a partilha de documentos e dados, e realizar pesquisas de informação de uma maneira mais fácil.

2

<http://www.w3.org/RDF/>

3

<http://www.w3.org/TR/owl-features/>

Este *working paper* irá endereçar a importância que esta recente tecnologia vem adquirindo no melhoramento dos processos das organizações. Para além disso, será efectuada uma visão geral sobre os projectos de investigação que decorrem, neste momento, sobre a Web Semântica, reforçando a importância que esta tecnologia vê adquirindo no panorama científico mundial.

Para tal, encontra-se dividido em algumas secções, na primeira serão introduzidos alguns conceitos sobre a Web Semântica, na secção seguinte é realçada a importância desta tecnologia nas organizações, ilustrada com alguns casos de estudo, posteriormente será efectuada uma referência aos projectos a decorrer nesta temática, bem como a apresentação de alguns dados estatísticos que consideramos interessantes. No final são tecidas algumas conclusões.

2. Web Semântica

2.1 O que é a Web Semântica?

Os seres humanos são capazes de usar Web para realizar determinadas tarefas como, encontrar a palavra inglesa para ‘casa’, reservar um livro da biblioteca, ou procurar por um preço reduzido para um DVD. No entanto, um computador não pode realizar estas tarefas sem a interacção humana, pois as páginas Web são concebidas apenas para serem lidas por pessoas, e não pelos sistemas computacionais. A Web Semântica é uma visão, onde a informação pode realmente ser compreendida pelos computadores, para que as máquinas possam realizar o trabalho tedioso envolvido na busca, partilha e combinação de informação na rede.

Tim Berners-Lee expressou em 1999, originalmente, a visão da Web Semântica como:

“ Eu tenho um sonho para a Web [em que os computadores] se irão tornar capazes de analisar todos os dados na rede – o conteúdo, *links* e transacções entre pessoas e computadores. A Web Semântica, que tornará isto possível, ainda não surgiu, mas quando isso acontecer, o dia-a-dia dos mecanismos de comércio, a burocracia e as nossas vidas diárias serão manipulados por máquinas a falar somente com outras máquinas. “

Neste momento, a Web Semântica compreende um conjunto de princípios de design, grupos de trabalho colaborativo, e uma variedade de tecnologias de base. Algumas delas incluem Resource Description Framework (RDF), RDF Schema (RDFS) e Web Ontology Language (OWL), os quais são destinados a fornecer uma descrição formal de conceitos, termos e relacionamentos dentro de um determinado domínio do conhecimento.

2.2 Tecnologias de Base

2.2.1 Resource Description Framework (RDF)

RDF é uma linguagem para representar informação na Internet. Ficheiros RDF são modelos ou fontes dados, vulgarmente conhecidos também por metadados, e que têm como principal objectivo criar um modelo simples de dados, com uma semântica formal.

Possuem três componentes básicas: *resources*, *properties*, e *statements*. Que tornam a linguagem altamente escalável.

2.2.2. Ontology Web Language (OWL)

A OWL (*Web Ontology Language*) é uma linguagem para definir e instanciar [ontologias](#) na [Web](#). Uma ontologia OWL pode incluir descrições de classes e suas respectivas propriedades e seus relacionamentos. A linguagem OWL foi projectada para o uso por parte de aplicações que necessitam processar o conteúdo da informação, ao invés de apenas apresentá-la. Facilita a possibilidade de interpretação por máquinas do conteúdo da Web do que [XML](#), [RDF](#) e [RDFS](#), pois fornecer vocabulário adicional com uma semântica formal. A OWL foi baseada nas linguagens [OIL](#) e [DAML+OIL](#), e é hoje uma recomendação da [W3C](#).

OWL é vista como uma tecnologia importante para a futura implementação da [Web Semântica](#). Ela vem ocupando um papel importante em um número cada vez maior de aplicações, e vem sendo foco de pesquisa para ferramentas, técnicas de inferências, fundamentos formais e extensões de linguagem.

2.3 Ontologias

Na ciência da computação e na ciência de informação, uma ontologia é uma representação formal de um conjunto de conceitos dentro de um domínio e das suas respectivas relações entre esses conceitos. Em teoria, uma ontologia é “uma especificação formal explícita de uma conceitualização compartilhada”, que fornece o vocabulário comum, e que pode ser usada para modelar um domínio – isto é definir o tipo de objectos e / ou conceitos existentes, e suas propriedades e relações.

As ontologias são bastante utilizadas na Web Semântica.

2.4 Taxonomias

Uma taxonomia, ou esquema taxonómico, é uma classificação especial, disposta em uma estrutura hierárquica. Normalmente, é organizado por relações supertipo e subtipo, também chamado de generalização de relações de especialização. Neste tipo de relações, o subtipo, por definição, tem as mesmas propriedades, comportamentos e restrições que o supertipo mais algumas outras propriedades.

3.A Web Semântica nas Organizações

3.1 A importância da Web Semântica nas Organizações.

Hoje, a integração é uma das grandes prioridades das empresas europeias e mundiais. A comunidade Europeia está a investir, através do Sétimo Programa Quadro, mais de 200 milhões € em pesquisa envolvendo interoperabilidade entre empresas e semântica. A maioria das organizações já percebeu, que o uso das tecnologias associadas à Web Semântica são uma solução promissora para apoiar a cooperação organizacional entre as PME (Pequenas e Médias Empresas) que operam em ambientes de trabalho totalmente dinâmicos. As tecnologias de Web Semântica são cada vez mais consideradas como uma tecnologia chave para resolver os problemas de interoperabilidade e integração num mundo totalmente heterogéneo em respeito à natureza dos componentes, normas, formatos de dados, protocolos, etc. Além disso, acredita-se que a Web Semântica pode facilitar, não só a descoberta de componentes heterogéneos e a integração dos dados, mas também a comunicação entre indivíduos.

A interoperabilidade inter-empresarial é a chave para a implementação de uma ideia de uma economia baseada no conhecimento, onde as redes de empresas (PME, em particular) podem obter vantagens das peculiaridades do tecido produtivo europeu. As novas tecnologias de interoperabilidade, dentro e entre as empresas, necessitam emergir radicalmente para resolver um problema que não tem sido suficientemente abordado pela comunidade científica. De facto, gerir a semântica de interacção dos negócios é uma tarefa desafiadora de integração, porém a Web Semântica tem o potencial de mitigar estes problemas.

Em suma, podemos afirmar que os princípios de tecnologia e design da Web Semântica fornecem uma poderosa *framework* que torna possível às empresas, continuar a atingir os objectos de negócio em face da mudança e da crescente dependência do uso efectivo da informação no processo empresarial.

3.2 Desafios

As empresas disponibilizam um vasto conjunto de informações para muitas pessoas no mundo: *blogs*, *wikis*, sites na web, etc. Isto desencadeia inúmeros recursos e possibilidades. Porém o volume da informação que está a ser gerada e disponibilizada é enorme, e os sistemas de software actuais não estão habilitados para lidar com isso. Actualmente, as empresas não estão preparadas, para ambientes onde tanta informação é gerada todos os anos. Uma vez que o sucesso do negócio depende muita das vezes, de o tomador de decisão ter acesso atempado a uma correcta informação e conhecimento, surge a necessidade de que existam sistemas que sejam capazes de fornecer esse conhecimento. É portanto firme a convicção de que, dado ao crescente papel da Internet na actividade empresarial e do puro crescimento do volume de informação disponível, haverá uma crescente demanda por tecnologias associadas à Web Semântica.

3.3 Casos de Uso

Neste capítulo descrevemos alguns casos de uso das tecnologias de Web Semântica dentro das organizações. Os primeiros exemplos descrevem-nos como a Web Semântica poderia desempenhar um importante papel dentro de uma organização de serviços e de manufactura. O caso de uso seguinte descreve a implantação de tecnologias de Web Semântica numa empresa de ciências da vida. O caso de uso final destaca o uso da Web Semântica para a pesquisa de conteúdo dentro de uma empresa tecnológica. O capítulo é finalizado com uma breve descrição de algumas outras áreas de aplicação, onde se espera que haverá um benefício significativo na adopção de tecnologias associadas à Web Semântica.

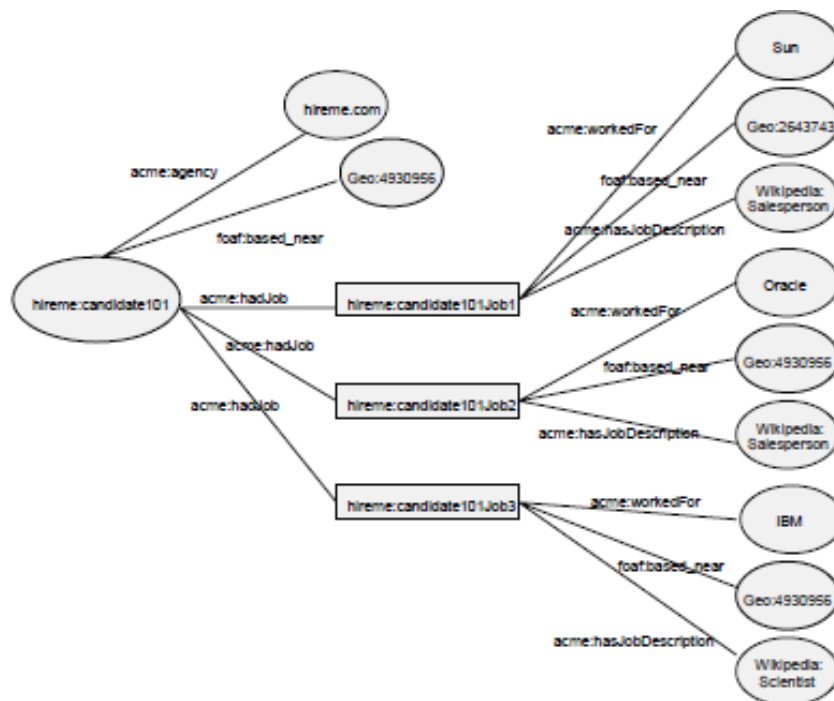
3.3.1 Melhorar a Eficácia dos Serviços de Recrutamento

Normalmente, quando uma empresa está a pensar contratar uma pessoa com um determinado conjunto de certificações e aptidões, inicia um processo de pesquisa, junto das empresas de recrutamento que trabalham com a respectiva empresa. Cada agência de recrutamento irá então, realizar uma pesquisa à sua base de dados para identificar os

melhores candidatos possíveis para a posição. Uma vez identificados os candidatos, a própria empresa é que terá que compilar e gerir todas as listas respectivas de candidatos. Seria bem mais eficiente, se a empresa que está a recrutar, fosse capaz de realizar uma única consulta em todas as bases de dados das empresas de recrutamento, obtendo assim, toda a informação num formato consistente. Isto permitiria, para além de poupar tempo na descrição da candidatura às agências de recrutamento, minimizar totalmente os dissentidos providentes do processo de comunicação entre as agências e a empresa. Adicionalmente, permitiria identificar directamente os candidatos que a empresa considera serem os ideais para a candidatura.

A fim de permitir a consulta de várias bases de dados, de entre as várias empresas de recrutamento, a empresa contratante deveria incentivar todas as agências a disponibilizar um subconjunto da sua informação em RDF, utilizando um vocabulário comum e expô-la através da ferramenta SPARQL.

Figura 1: Possível esquema para representar informações sobre uma candidatura individual.



Depois de todas as empresas de recrutamento disponibilizarem a sua informação em RDF através de um *endpoint* SPARQL, será simples para a empresa consultar os

candidatos em todas as agências. É bastante simples também para a empresa integrar uma nova agência na sua lista de parceiros.

Em forma de melhoramento, existem inúmeras razões pelas quais a empresa pode decidir prorrogar a sua infra-estrutura de Web Semântica para incluir OWL. Por exemplo, seria bastante útil se houvesse uma classe que descreve-se candidatos com conhecimentos jurídicos, evitando assim a necessidade de procurar candidatos com títulos de trabalho que incluem advogado, solicitador ou representante legal.

3.3.2 Melhorar a manufactura ágil

Os fabricantes de alimentos operam em ambientes que existe uma pressão severa em relação ao custo. São das poucas indústrias em que as decisões de compra são feitas com base em diferenças de preços de alguns tostões. Esta influência dos custos coloca uma pressão elevada sobre os fabricantes para produzir o produto certo, na quantidade certa, e no momento certo. Isto requer uma melhor compreensão da procura do consumidor e da habilidade de rapidamente obter ingredientes dos seus fornecedores. É também uma indústria que necessita de responder perante as várias entidades reguladoras. A fim de controlar eficazmente estes factores de negócio, as empresas utilizam um grande número de aplicações corporativas. Uma das principais aplicações tem sido a utilização de software de planeamento de recursos, que é projectado para lidar com vendas, gerência de pedidos, controlo de stocks, facturação e outras funções de escritório. Outra aplicação crítica é o sistema de execução de fabricação. Este software permite analisar o desempenho de forma dinâmica e determinar o factor de produção, entre outros parâmetros. Estas aplicações têm evoluído bastante nos últimos anos, no entanto, mesmo com estes melhoramentos, é frequentemente difícil integrar estas diferentes aplicações e muitas vezes suportar os diferentes modelos de dados subjacentes.

A Web Semântica fornece a capacidade de integrar todos estes conceitos numa ontologia. Em termos de arquitectura, a ontologia é utilizada como a interface entre a aplicação e a base de dados. Com esta abordagem, o foco da base de dados é apenas o armazenamento persistente da informação. Em termos da aplicação, o foco é dado à aplicação na lógica empresarial, que faz uso, dos termos definidos pela ontologia,

criando assim um modelo unificado de dados através de múltiplas bases de dados. Os benefícios desta abordagem, incluem o facto de que as ontologias mapeiam e organizam os conceitos chave de uma organização, que têm uma tendência de mudar de forma relativamente lenta. A engenharia de software é, portanto, mais robusta contra as alterações quando as ontologias são utilizadas. Tendo as definições dos conceitos, restrições e relações na ontologia definidas, torna mais fácil a reutilização dos dados por parte das aplicações para as quais não foram originalmente destinadas. É também mais acessível para as pessoas que entendem do negócio ajudar no desenvolvimento de ontologias, do que na criação de esquemas das bases de dados. Além disso, é relativamente mais fácil estender as capacidades de uma ontologia, adicionando relações e restrições, ou ligando diferentes ontologias. Por exemplo, existiu uma regulamentação nos Estados Unidos, que imponha um maior controlo nas etiquetas dos produtos que contêm ingredientes alérgicos comuns. Ao adicionar novas definições de classes de produto, é possível uma classificação automatizada que torna fácil a identificação destes produtos. Similarmente, para apoiar a comercialização, podem ser definidas novas classes, que satisfaçam por exemplo as últimas dietas populares de perda de peso.

Tirando partido das tecnologias da Web Semântica, as empresas podem incorporar novos dados que sejam considerados relevantes, de uma forma simples.

3.3.3 Integração de dados científicos heterogéneos

Muitas empresas farmacêuticas estão interessadas nos recursos prometidos pela Web Semântica, principalmente na integração dos dados. Isto ocorre porque a descoberta e desenvolvimento de medicamentos é um processo bastante demorado e caro. É também importante para as empresas estar ciente de ofertas competitivas ou patentes que possam reduzir o potencial do mercado. A fim de ser capaz de tomar estas decisões, as empresas farmacêuticas necessitam ter uma visão integrada dos dados. A integração dos dados, no entanto, tem-se provado ser um processo bastante complexo. Os dados são gerados em diferentes departamentos, onde diferentes terminologias são utilizadas. Para além disso, os dados em si são bastante heterogéneos e consistem em vários tipos de dados que incluem várias estruturas químicas, sequências biológicas, imagens, caminhos biológicos, etc. Muitas empresas têm tentado criar *data warehouses* que

contêm todos esses dados, mas muitos encontraram nesta abordagem uma enorme falta de flexibilidade, necessária dentro de uma disciplina científica. Consequentemente, as empresas estão a explorar abordagens alternativas para a integração dos dados. Actualmente, um grande número de companhias farmacêuticas está a explorar o uso de tecnologias relacionadas com a Web Semântica como *framework* para a integração heterogénea dos dados. A maioria dos projectos está focada apenas na integração de dados relacionados com novas descobertas de medicamentos, embora em última análise, algumas empresas estejam a se esforçar para uma fornecer uma visão unificada dos dados desde o laboratório até ao período de observações clínicas.

As implementações relacionadas com a Web Semântica nas empresas farmacêuticas têm desenvolvido ferramentas que permitem compreender as principais entidades encontradas dentro do processo de descoberta de novos medicamentos. Estas entidades normalmente incluem genes, proteínas, compostos, amostras, doenças, projectos e empresas. Quando um utilizador interroga uma instância de uma entidade, torna-se possível ver todas as informações que se relacionam com o objecto. Frequentemente, os resultados de tais consultas mostram um gráfico com todas as entidades, de modo a facilitar a visualização e navegação das relações entre as entidades. Esta abordagem permite aos cientistas descobrir informação à medida que navegam, através do conhecimento disponível, sem ter necessariamente uma consulta específica em mente. Quando as ontologias são usadas em combinação com esta ferramenta, permitem ao utilizador obter toda a informação de interesse, mesmo que uma outra terminologia tenha sido usada para gravar os dados. Nas consultas, pode também ajudar a reconhecer termos idênticos, usados em diferentes contextos.

Existem uma série de outras áreas, nas quais a indústria farmacêutica está interessada em continuar a explorar o uso da Web Semântica. Uma área de interesse adicional está em identificar se um caminho comum biológico, se relaciona com duas observações que ocorrem separadamente, por exemplo, uma alteração na patologia do cérebro e a mudança da temperatura do corpo. Isto pode ser efectuado através da integração de uma ontologia que relaciona as condições clínicas com sintomas, com outra ontologia que relaciona as condições clínicas com caminhos biológicos.

3.3.4 Optimização da procura e navegação em uma empresa

A pesquisa de conteúdo é uma área em que várias empresas estão a usar a Web Semântica. O caso de uso seguinte, descreve como as tecnologias de Web Semântica têm sido aproveitadas para melhorar a pesquisa de uma grande empresa. A *Oracle Technology Network (OTN)*, parte integrante do site da empresa, é a principal fonte de informações técnicas da comunidade Oracle. O site fornece acesso a várias documentações de produtos, notificações de lançamentos de produtos, software, *blogs*, *podcasts*, e um fórum de discussão. A riqueza, complexidade e dinamismo destas informações tornam-se um desafio para a pesquisa tradicional e sobre a navegabilidade para fornecer o acesso efectivo à informação. A *Oracle* encontra-se a trabalhar com a *Siderean Software* para aplicar as tecnologias baseadas em Web Semântica no site, em ordem de resolver este problema. O site encontra-se disponível em <http://otnsemanticweb.oracle.com>.

Esta combinação de tecnologias melhora o acesso à informação através da agregação de várias fontes de conteúdo e seu respectivo fornecimento através de um único portal. A aplicação permite uma abordagem comum para pesquisas e consultas, este processo unificado de acesso à informação é bastante valioso, pois permite aos utilizadores identificar todas as informações de interesse com um único pedido. Os utilizadores têm também a capacidade de personalizar o seu ambiente, seleccionando os itens que querem incorporar e o modo de apresentação da informação. Para manter o site actualizado, a cada hora a aplicação ‘puxa’ conteúdo de interesse para a *Oracle Developer Community* através de vários *feeds RSS*. Apesar de cada *RSS* conter alguns metadados que o descreve, muitas vezes esta informação não é suficientemente detalhada ou organizada. O *Siderean Metadata Assembly Platform* reforça estes metadados usando técnicas de correspondência para identificar o assunto de cada objecto. Neste ponto, XSLT é usado para converter o RSS baseado em XML para sintaxe RDF. Esta abordagem permite o suporte necessário para uma experiência rica de pesquisa, com base na navegação dinâmica das entidades e dos seus relacionamentos. A interface de pesquisa fornece também a capacidade de criar *RSS* personalizados, baseado em consultas definidas pelo utilizador. Como os conceitos e as entidades

presentes nos metadados se encontram padronizados, os *feeds* construídos usando esta tecnologia tendem a ser mais segmentados.

Esta solução de pesquisa foi implementada por uma empresa que quer, substancialmente, aumentar a qualidade da pesquisa e de navegação dos seus clientes. Uma extensão interessante para esta capacidade seria a de alargar a pesquisa e navegação para incluir aplicativos de negócio. Para alcançar este objectivo, seria necessário estender a ontologia para incluir todas estas entidades representadas no âmbito das aplicações de negócios.

3.3.5 Áreas Adicionais da Web Semântica

A secção anterior destacou alguns casos de uso da Web Semântica. Porém, existem outras áreas de aplicação onde estas tecnologias podem fornecer benefícios significativos. As áreas são:

- **Financeira:** O domínio financeiro, onde a informação é bastante valiosa e grandes volumes de informação são trocados diariamente, naturalmente, pode beneficiar da utilização de modelos de domínio explícito, compartilhados por todos os intervenientes nos diferentes mercados financeiros.
- **Governamental:** Motores de busca inteligentes, baseados no uso intensivo da Web Semântica, aumentam significativamente a compreensão do computador perante consultas de um cidadão em termos dos serviços disponíveis online. Este entendimento permite melhorar o desempenho e cobertura dos serviços disponíveis, bem como os índices de satisfação de qualquer utente, face aos serviços disponibilizados pelas instituições públicas.

4. Projectos de Investigação na Web Semântica

Em termos de projectos de investigação, encaramos este capítulo como a secção do futuro da Web Semântica, isto é, após apresentadas as possíveis aplicações da Web Semântica nas organizações, apresentamos um conjunto de projectos de investigação, que se encontram a decorrer e que representam as inovações nesta tecnologia.

Para obter informação sobre estes projectos, a maioria iniciados neste ano de 2009, foi consultado o Programa do Sétimo Quadro⁴ (*7th Framework Programme*), que abrange o período de 2007 a 2013 e que tem como principal objectivo, permitir à União Europeia colocar a sua política de investigação à altura das suas ambições económicas e sociais, através do Espaço Europeu de Investigação (EEI). Apesar deste programa englobar toda a União Europeia, e desse modo, Portugal incluído, foi nossa opção consultar a Fundação para a Ciência e a Tecnologia⁵ (FCT) de modo a obter resultados em relação a projectos Portugueses, para conseguir perceber de um modo mais eficaz qual o tipo de investigação está a decorrer.

Portanto, existem dois principais objectivos nesta secção de investigação: perceber quais os projectos a nível Europeu que estão agora a iniciar, bem como aqueles que especificamente decorrem em Portugal.

Existe bastante informação neste capítulo, sobretudo obtidas através das equipas intervenientes em cada projecto de investigação, e que incidem especificamente sobre os subsídios atribuídos a cada um dos projectos, e sobre todas as informações relativas ao âmbito, contexto e objectivo de cada investigação.

Todos estes projectos na área da Web Semântica, podem ser vistos como inovadores, na medida que esta é uma área bastante recente, principalmente na sua aplicação a organizações, pois em termos de investigação faz já parte do panorama científico mundial a algum tempo. Para cada projecto, pretendemos especificar alguns dados, mas mais especificamente o estudo irá centrar-se nos principais intervenientes de cada

4

<http://cordis.europa.eu/fp7/>

5

<http://www.fct.mctes.pt/>

projecto, de modo a perceber qual a influência de cada universidade e país, para a investigação na área da Web Semântica, permitindo-nos assim no final deste capítulo, efectuar algumas conclusões interessantes. Sobretudo interessa-nos perceber quais os países mais influentes na área da investigação com tecnologias relacionadas com a Web Semântica. Não pretendemos de modo algum, fazer um resumo de cada projecto em termos de conteúdo, embora seja importante perceber realmente qual o objectivo do próprio e da investigação propostas. O *focus* primordial será dado aos factores adjacentes ao próprio âmbito do projecto, tais como pessoas e organizações intervenientes, e respectivo financiamento. Obviamente existem muitos projectos, no Programa do Sétimo Quadro foram encontrados dezassete relativos a Web Semântica, embora uns com mais relevância do que outros. A escolha dos projectos incluídos neste *working paper*, teve como elemento decisivo, a relevância dos seus conteúdos no âmbito concreto da Web Semântica e da sua importância nas organizações. De seguida, são apresentados com algum detalhe os vários projectos.

4.1 Semantic Oriented Multimedia Indexation and Retrieval

O objectivo deste projecto é fornecer uma solução para a obtenção eficiente de informação orientada à semântica, através do suporte de indexação tradicional de multimédia, de técnicas de classificação, e do uso das tecnologias da Web Semântica moderna. O principal desafio proposto pela equipa do projecto é solucionar o problema de ligar a “falha” semântica que existe entre as características multimédia de baixo nível e os conceitos de alto nível, que descrevem o conteúdo multimédia. Podem ser consideradas características multimédia de baixo nível: a cor, textura, luminosidade e o próprio som. Os conceitos de alto nível, basicamente são informações textuais que vem agregada ao próprio conteúdo multimédia, por exemplo, o uso de *tags* para descrever o seu conteúdo.

No âmbito das organizações, uma estação de televisão, que contém e gere grandes conteúdos multimédia, pode utilizar esta aplicação que indexa e obtém conteúdos multimédia. Algumas informações:

Início: 2009-05-01

Conclusão: 2011-04-30

Estado: Activo

Financiamento Atribuído: 220.594 euros

Entidade proponente: Universite Paul Sabatier Toulouse III

4.2 Large scale semantic computing semantic Web technologies distributed reasoning probabilistic reasoning web-scale inference information retrieval (LARKC)

Este projecto ⁶ tem como objectivo construir uma aplicação para lidar com grandes quantidades de conhecimento, ou seja, criar uma plataforma para o raciocínio massivo distribuído, que irá eliminar todas as barreiras ligadas à escalabilidade, mesmo em relação à obtenção de informação distribuída. Basicamente, o objectivo desta aplicação é distribuir informação e métodos de obtenção da mesma, bem como a sua classificação, para contornar o problema da escalabilidade. Existem várias máquinas, com diferentes informações e métodos. O objectivo da aplicação é conseguir utilizar todo este hardware distinto, para confrontar a informação e métodos de cada máquina de modo a conseguir obter uma informação fiável e esclarecedora. Para algumas organizações, este projecto poderá vir a ser interessante, dado que existem muitas empresas especializadas em guardar informação confidencial de outras. A obtenção de informação (*information retrieval*) é algo muito relevante para muitas empresas, e cada vez mais se desenvolvem processos para melhorar esses tipos de técnicas. A equipa de investigadores do projecto é composta por um grupo de investigadores na área de *Computer Science*, *Web Science* e *Cognitive Science*, sendo alvos o sector das telecomunicações e da área da ciência viva. Os intervenientes deste projecto, constituem um total de doze países diferentes, o que de facto representa um factor importante em termos de financiamento. Alguma informação relativa ao projecto:

Início: 2008-04-01

Conclusão: 2011-09-30

Estado: Activo

Financiamento Atribuído: 7.5 milhões de euros

Entidade proponente: Universitaet Innsbruck - Digital Enterprise Research Institute

Em termos de participantes no projecto, existem dezassete parceiros:

- CYCORP INC. (Estados Unidos)
- SEMANTIC TECHNOLOGY INSTITUTE INTERNATIONAL (Áustria)
- ASTRAZENECA AB (Suécia)
- CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O. (Eslovénia)
- SALTUX INCORPORATED (República da Coreia)
- CEFRIEL - SOCIETA CONSORTILE A RESPONSABILITA LIMITATA (Itália)
- UNIVERSITAET STUTTGART (Alemanha)
- MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V. (Alemanha)
- CENTRE INTERNATIONAL DE RECHERCHE SUR LE CANCER (França)
- BEIJING UNIVERSITY OF TECHNOLOGY (China)
- VERENIGING VOOR CHRISTELIJK HOGER ONDERWIJS WETENSCHAPPELIJK ONDERZOEK EN PATIENTENZORG (Holanda)
- INFORMATION RETRIEVAL FACILITY SOCIETY (Áustria)
- THE UNIVERSITY OF SHEFFIELD (Reino Unido)
- SIRMA SOLUTIONS JSC (Bulgária)
- ONTOTEXT AD (Bulgária)
- POLITECNICO DI MILANO (Itália)
- SIEMENS AG (Alemanha)

Como podemos observar acima, existem entidades para representar todas as áreas de conhecimento da investigação.

4.3 Knowledge in a Wiki (KIWI)

O objectivo deste projecto⁷ é desenvolver um sistema avançado de gerenciamento de conhecimento (“KIWI system”), baseado numa *wiki* semântica. Basicamente, este sistema irá suportar a criação e partilha de conhecimento colaborativo, usando descrições semânticas, bem como raciocínio. Este projecto propõe uma nova abordagem de conhecimento, seguindo a filosofia das *wikis*, mas utilizando as potencialidades dos métodos da Web Semântica, permitindo assim um melhor desempenho de quem usa este tipo de ferramentas para gerir informação relevante. Um dos casos de estudo deste projecto é a empresa *Sun*, que para o desenvolvimento de um determinado tipo de

software, utiliza equipas bastante vastas. O objectivo da KIWI, em particular na empresa *Sun*, é tentar melhorar, em termos de eficiência, o desenvolvimento do software, pois a partilha/obtenção de informação é feita de uma forma mais eficaz.

Os principais resultados do projecto serão: (1) uma melhoria em relação à visão do conceito de *wiki*, combinando o conceito de *wiki* com tecnologias semânticas, levando a uma gestão flexível do raciocínio orientado a problemas. (2) Um ambiente colaborativo, *web-based* que fornece suporte para a partilha, criação e coordenação de conhecimento em desenvolvimento de software e em projectos de gestão de conhecimento. (3) A avaliação deste sistema em dois casos de uso específicos na indústria.

Como já foi especificado acima (através do caso de estudo da *Sun*) este projecto tem um grande grau de aplicabilidade nas organizações, de modo a melhorar o seu funcionamento e a aumentar a qualidade da informação que nela circula. Trata-se de um projecto de extrema relevância para todas as organizações. Algumas informações acerca do projecto:

Início: 2008-03-01

Conclusão: 2011-02-28

Estado: Activo

Financiamento Atribuído: 2.69 milhões de euros

Entidade proponente: [Salzburg Research Forschungsgesellschaft mbH](#)

Em termos de participantes no projecto, existem seis parceiros:

- SUN MICROSYSTEMS CZECH SRO (República Checa)
- SEMANTIC WEB COMPANY BLUMAUER & PARTNER OG (Áustria)
- AALBORG UNIVERSITET (Dinamarca)
- LOGICA DANMARK A/S (Dinamarca)
- LUDWIG-MAXIMILIANS-UNIVERSITAET MUENCHEN (Alemanha)
- VYSOKE UCENI TECHNICKE V BRNE (República Checa)

4.4 Collaborative annotation of a large biomedical corpus (CALBC)

O grande propósito deste projecto, é copular a problemática dos investigadores na área do *text mining* e biomédica, às questões de anotação semântica de corpus extensos de

texto a conjuntos semânticos igualmente extensos. O que este projecto propõe como resultado, é basicamente um conjunto de textos da área da biomédica, devidamente anotados em relação a entidades mencionadas. Entidades mencionadas são entidades que surgem numa determinada frase, num texto. Essas entidades encontram-se inseridas num determinado contexto, de modo a serem semanticamente classificadas, ou seja, uma entidade mencionada só é correctamente classificada desde que o seu contexto seja tomado em conta, através de evidências internas e externas. Como resultado do projecto, obtém-se um conjunto de textos anotados, com entidades mencionadas identificadas e semanticamente classificadas. Esses textos anotados, serão bastante úteis a toda a comunidade biomédica, pois permite melhorar o conhecimento nesta área, bem como melhorar as técnicas de classificação, muito utilizadas na biomédica. Algumas informações acerca do projecto:

Início: 2009-01-01

Conclusão: 2011-06-30

Estado: Activo

Financiamento Atribuído: 2.2 milhões de euros

Entidade proponente: European Molecular Biology Laboratory Administration Department

Em termos de participantes no projecto, existem três parceiros:

- LINGUAMATICS LIMITED (Reino Unido)
- ERASMUS UNIVERSITAIR MEDISCH CENTRUM ROTTERDAM (Holanda)
- FRIEDRICH-SCHILLER-UNIVERSITAET JENA (Alemanha)

4.5 OntoWiki semantic collaboration for enterprise knowledge management, e-learning and e-tourism (ONTOWIKI)

Este projecto ⁸ tem como principal objectivo utilizar a ferramenta já existente OntoWiki, e explorar a sua aplicação a organizações, mais especificamente em três alvos diferentes: Enterprise Knowledge Management, *E-learning* e *E-tourism*. Este projecto consiste em iniciar um processo de *setup* no tipo de organizações referidas, visto que esta ferramenta já existe e é *open-source*. Este projecto é de extrema relevância para as organizações, pois consiste numa adaptação de um tipo de software que permitirá um

8 <http://ontowiki.eu/>

melhor rendimento em termos do domínio específico de cada empresa. Abaixo, encontram-se algumas informações acerca deste projecto:

Início: 2008-11-01

Conclusão: 2010-10-31

Estado: Activo

Financiamento Atribuído: 943.800 euros

Entidade proponente: Universitaet Leipzig

Em termos de participantes no projecto, existem seis parceiros:

- PUNKT. BLUMAUER KALTENBOCK KOLLER OEG (Áustria)
- INSTITUT FÜR ANGEWANDTE INFORMATIK EV (Alemanha)
- B2 IZOBRAZEVANJE IN INFORMACIJSKE STORITVE D.O.O (Eslovénia)
- BI BUSINESS INTELLIGENCE GMBH (Alemanha)
- OPENLINK GROUP LIMITED (Reino Unido)
- VAKANTIELAND (Holanda)

4.6 CASPOE – Caracterização Semântica e Pragmática de Objectos Educativos

Este projecto CASPOE ⁹ é financiado pela Fundação para a Ciência e a Tecnologia, e tem como principal objectivo criar um protótipo de um modelo de um repositório de objectos educativos, no sentido em que estes possam ser utilizados para ensino e aprendizagem. Basicamente, podemos dizer que um repositório de objectos educativos, é um repositório de informação que pode ser usada para fins da educação. Para os participantes deste projecto, o uso dos repositórios educativos ainda não está suficientemente desenvolvido, daí a proposta deste modelo, que faz uso da Web Semântica para contextualizar os objectos, de modo a criar um repositório que caracterize todo o material disponível, para serem aplicados a métodos de aprendizagem. Este modelo é independente da área em que actua, ou seja, através da especificação do domínio no modelo, este pode funcionar para qualquer área de conhecimento. Basicamente criam-se ontologias de domínio, que servem para relacionar

todos os conceitos e relações entre eles, sem qualquer estratégia pedagógica, ou seja, o uso das ontologias é somente útil para estabelecer relações entre os conceitos presentes do repositório. Para cada objectivo educativo, é construída uma taxonomia, que utiliza a ontologia de domínio já definida. Existem ainda outra componente do modelo que se baseia em ontologias: Ontologia de *Learning Design de Aprendizagem*. Relativamente ao conceito de *Learning Design*, este está relacionado com a aprendizagem, no sentido em que existem vários padrões e cenários possíveis para aprender, permitindo assim a sistematização dos vários processos de aprendizagem. No caso de *e-learning* seria aprendizagem online. Todas as entidades importantes contidas neste projecto, serão representadas por meio de ontologias, sendo a Web semântica uma forte componente neste projecto.

Pensado agora neste projecto, aplicando-o às organizações, este repositório de objectivos educativos poderá ser encarado como um sistema de formação inteligente para os funcionários de uma determinada empresa. Dadas as características do modelo a ser desenvolvido, seria possível numa dada organização, este mesmo sistema inteligente de formação possuir diversas áreas de conhecimento, consoante o alvo a atingir. Deste modo, encaramos este projecto como extremamente interessante, dado à sua aplicabilidade nas organizações.

De seguida, iremos apresentar algumas informações relativamente a financiamento e datas do projecto:

Início: 2007-11-1

Conclusão: 2010-10-31

Estado: Activo

Financiamento Atribuído: 93.800,00 Euros

Em termos de participantes no projecto, existem dois parceiros:

- **Instituto Superior de Engenharia do Porto (ISEP/IPP)**
- *Faculdade de Engenharia da Universidade do Porto (FEUP)*

O Instituto Superior de Engenharia do Porto, é a entidade proponente deste projecto.

4.7 SemanticShare – Ferramentas de Recursos para o Processamento Semântico

O principal objectivo deste projecto ¹⁰é o desenvolvimento de corpus anotados para língua portuguesa. Mais especificamente, para cada corpus, pretendem-se agregar uma série de informações que possam ser úteis para esses corpus anotados serem usados como bases empíricas para estudo da língua portuguesa, para servirem de conjuntos de treino para classificadores de português, ou para servirem de base para avaliar ferramentas de reconhecimento de língua portuguesa. Este projecto está mais relacionado com a própria semântica, do que especificamente na área da Web Semântica. Algumas informações relativas ao projecto:

Início: 2009-02-1

Estado: Activo

Financiamento Atribuído: 159.513 Euros

Em termos de participantes no projecto, existem dois parceiros:

- Fundação da Faculdade de Ciências (FFC/FC/UL)
- Pontifícia Universidade Católica do Rio Grande do Sul - Faculdade de Informática (FACIN/PUCRS)

Sendo a Fundação da Faculdade de Ciências, a instituição proponente.

4.8 Informação Estatística

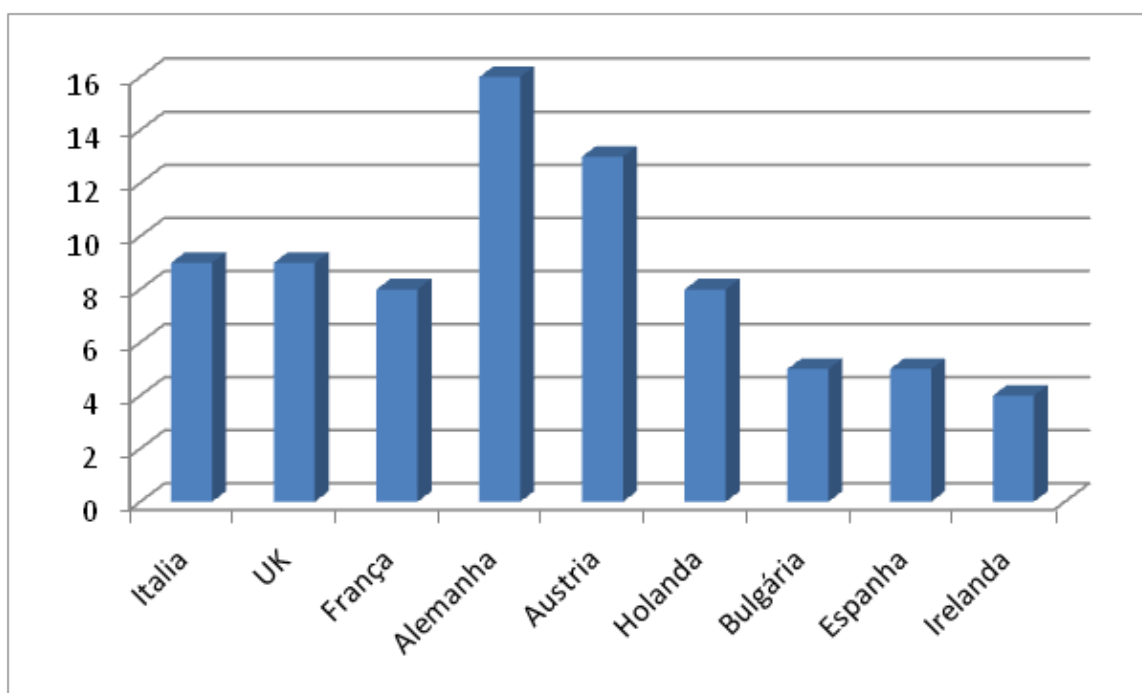
Apresentamos agora alguma informação estatística relativa aos intervenientes nos projectos relacionados com Web Semântica no Programa do Sétimo Quadro Europeu. Para isso, serão oferecidos alguns gráficos que nos permitem verificar quais os países que mais participam nos projectos, bem como aqueles que mais iniciativas possuem.

Participações por país

O gráfico seguinte, revela a presença de cada país nos dezassete projectos encontrados, relativos a Web Semântica:

Figura 2: Presença de cada país em função de cada projecto existente

Nº Presenças



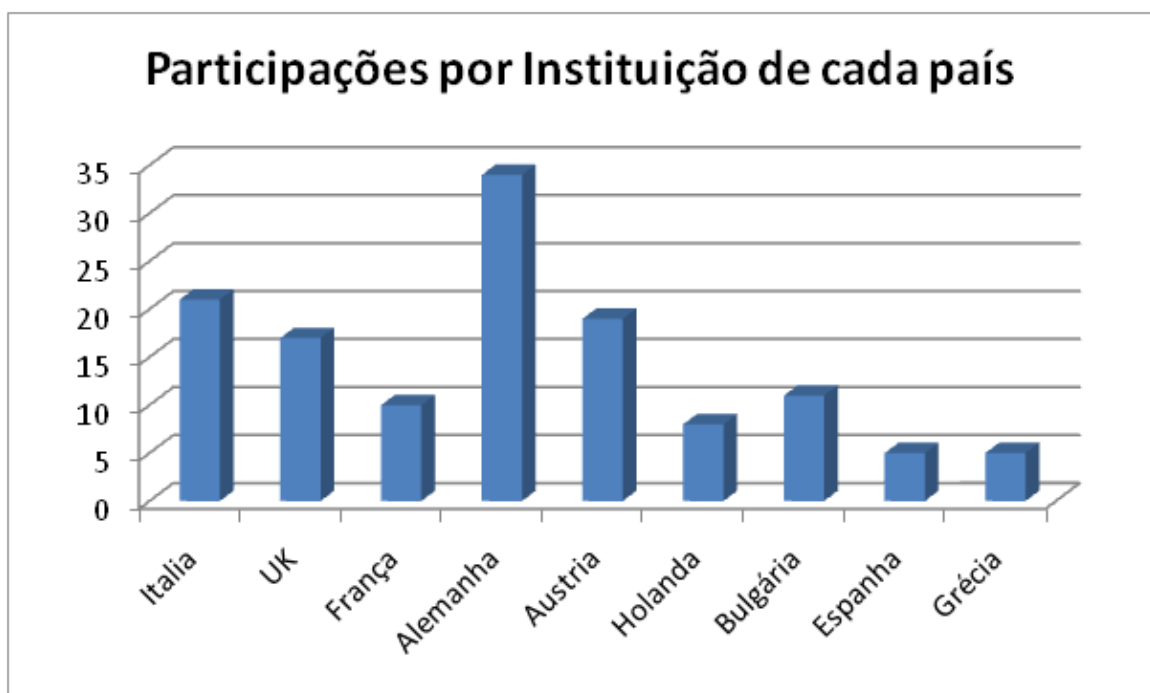
O número de países que se encontram no gráfico, não corresponde exactamente à totalidade dos países participantes, pois para os dezassete projectos que foram consultados, o número total de países distintos encontrados foi de vinte e sete. Para uma melhor compreensão do gráfico, os restantes países foram excluídos, embora alguns deles tenham mesmo participado em três dos dezassete projectos existentes (Finlândia e Eslovénia), sendo que os restantes tenham apenas participado uma ou duas

vezes (Suécia, República da Coreia, Roménia, Suíça, Turquia, Polónia e Grécia). Um facto que deve ser também realçado, é que Portugal não está incluído em nenhum dos projectos financiados pelo Programa do Sétimo Quadro, o que nos leva a concluir que estamos muito abaixo do esperado e pretendido em relação a esta tecnologia.

Embora este gráfico nos indique o número de presenças de cada país para cada um dos projectos consultados, não nos permite saber qual o número de presenças por instituição, obviamente ligada a um determinado país. O gráfico abaixo, mostra esses valores:

Figura 3: Presença de cada instituição de um país em função de cada projecto existente

Nº Presenças



Mais uma vez podemos observar que as instituições Alemãs, são sem dúvida, as mais participativas em actividades de investigação financiadas pelo Programa do Sétimo Quadro, seguidas das instituições da Itália e do Reino Unido. Verifica-se que, na maior parte dos projectos, existem mais do que uma instituição Alemã como interveniente. Concretamente, existe apenas um projecto, dos dezassete existentes, em que não entram intervenientes de nacionalidade Alemã.

Em jeito de conclusão, e com base na observação dos dois gráficos, podemos constatar que os países que mais participam em projectos de investimento financiados pelo Programa do Sétimo Quadro são: Alemanha, Áustria, Itália e Reino Unido. Sendo a Alemanha o país que mais se destaca, no desenvolvimento desta tecnologia.

Foram também observados todos os intervenientes, nos dezassete projectos do Programa do Sétimo Quadro, resultando num total de cento e vinte e sete participantes, nos quais se destacam:

- **Universitaet Innsbruck** (*University of Innsbruck*) – Universidade Pública localizada em *Innsbruck*, na Áustria.
- **SAP AG** – Uma multinacional de desenvolvimento de software e consultoria, sediada na Alemanha.
- **Institut national de recherche en informatique et en automatique**¹¹ (*National Institute for Research in Computer and Control Sciences*) – Instituição de investigação em França, focada essencialmente em ciências da computação, teoria de controlo e matemáticas aplicadas.
- **Sirma solutions JSC** – Um grupo de empresas de conhecimento, que fornecem consultoria em tecnologias de informação (teve origem na Bulgária).
- **Ontotext AD**¹² – Uma empresa pertencente ao grupo Sirma Solutions, com o principal *focus* no desenvolvimento e investigação de tecnologias de gestão de conhecimento, Web semântica e *Web Services*.
- **The Open University** – Universidade, fundada e financiada pela governo do Reino Unido, focada também em muitos projectos de investigação (cerca de 25 áreas distintas).
- **Semantic Technology Institute International**¹³ – Uma “rede global” com uma base muito forte em termos de desenvolvimento e investigação de aplicações relacionadas com a Web Semântica.

11 <http://www.inria.fr/>

12 <http://www.ontotext.com/>

13 <http://www.sti2.org/>

Relativamente a projectos de I&D na área de Web Semântica, em Portugal, não foram encontrados muitos com grande substancialidade. Os dois projectos aqui apresentados, são os únicos existentes desde o ano de 2006.

5. Conclusão

Nos últimos anos, muito progresso foi feito no desenvolvimento de ideias e ferramentas para permitir o crescimento da Web Semântica. O núcleo das tecnologias standards atingiram um nível de maturidade tal, que lhes permite ser amplamente adoptadas. Uma vasta gama de ferramentas *open-source* e software comercial encontram-se disponíveis para qualquer pessoa.

Há muitos benefícios que podem ser alcançados com a extensa adopção da visão da Web Semântica, por exemplo, é bastante mais simples para qualquer indivíduo encontrar informação de interesse na Web e efectuar operações sobre esses mesmos dados. No entanto, muitas implementações tecnológicas na área da Web Semântica, têm surgido no contexto das organizações, relevando-se muito importantes, nesta sociedade de informação, para a melhoria do processo empresarial e de negócio. No decorrer deste *working paper* foram apresentados alguns casos de uso que atestam as melhorias dos processos das empresas. Serão estas tecnologias que irão preparar o caminho para a criação de aplicações mais gerais no âmbito da Web Semântica.

Porém, a Web Semântica ainda têm alguma distância a percorrer antes de atingir um ponto de adopção generalizada, apesar de algum destaque que vêm adquirindo, na Europa, sobretudo na Alemanha, têm-se verificado um crescimento importante no número de projectos relacionados com esta tecnologia. Espera-se que, futuramente, em particular no Programa do Oitavo Quadro Europeu (*8th Framework Programme*), se encontre reflectida a enorme importância desta tecnologia, promitente assim da sua adopção generalizada.

6. Bibliografia

1. [Cardoso, Jorge; Martin Hepp, Miltiadis D. Lytras](#) (Eds.): *The Semantic Web: Real-World Applications from Industry. Semantic Web And Beyond Computing for Human Experience* Vol. 6 Springer 2007, ISBN 978-0-387-48531-7.
2. Antoniou, G. and van Harmelen, F. *A Semantic Web Primer*. Cambridge, MA; MIT Press, 2004.
3. Sheth, A. and Lytras, M., *Semantic Web-based Information Systems: State-of-the-art Applications*, 2006, IGI Global, ISBN: 1599044277.
4. Hepp, M., De Leenheer, P., de Moor, A. and Sure, Y., *Ontology Management: Semantic Web, Semantic Web Services, and Business Applications*, 2007, Springer, ISBN: 038769899X.
5. Wache, H., T. Voegelé, V. U., S. H., G. Schuster, H. Neumann and S. Huebner (2001). *Ontology-based integration of information - a survey of existing approaches*. IJCAI workshop on Ontologies and Information Sharing, Seattle, WA. USA. pp. 108–117.
6. Li, Y., S. Thompson, Z. Tan, N. Giles and H. Gharib (2003). *Beyond Ontology Construction; Ontology Services as Online Knowledge Sharing Communities*. The SemanticWeb - ISWC 2003. Berlin/Heidelberg, Springer. 2870/2003.
7. Stephens S.M., A. Morales, et al. (2006) *Application of Semantic Web Technology to Drug Safety Determination*. IEEE Intelligent Systems. 21: 82-86.
8. Shannon, P., A. Markiel, et al. (2003) *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Research 13: 2498-2504.
9. Davies, J., Studer, R., and Warren, P. *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. Chichester, UK; John Wiley & Sons, 2006.
10. Castells, P., Foncillas, B. and Lara, R. (2004). *Semantic Web Technologies for Economic and Financial Information Management*. In Proceedings of the 1st European Semantic Web Symposium (ESWS 2004), Heraklion, Greece, May 2004.